**Introduction**

*“The best way to predict the future is to create it”,* said Abraham Lincoln, without intending, he Formulated accurately the core spirit of each sportsman who tries to predict his future accomplishments. Nevertheless, when I use Machine Learning to analyze many observations of such motivated athletes, it is interesting to identify and measure the importance that certain characteristics have on the success of participating in sports events. Through this rationale and my passion for basketball, I decided to base my ML project on the most famous sport’s tournament for data scientists – the "March Madness".

The National Collegiate Athletic Association (NCAA) Men's Basketball Tournament is informally referred to as "March Madness".

This project is a clear challenge against all odds. there is a 64-team pool with 63 games to predict. Given the sporting nature of a basketball game, it is interesting to identify and measure the importance that certain characteristics influence the success of participating in the tournament. Despite being very difficult to reach great accuracies, people continue to research and try their best. Mathematically speaking, perfectly fill a March Madness bracket is one of the most unlikely things on sport’s events, with a chance of .

The problem is not classification of individual teams, but rather predicting the outcome of a match between any two teams. since I have the data of the last 25 years this data, I would try to use Machine Learning to find out what statistics most correlate with a team winning a match-up.

Our ML model will analyze information about two teams (Team 1 and Team 2) as input, and then output a probability of Team 1 winning that matchup.

By running this module over the first-round matches (that I get as an input), I can simulate the whole tournament until the final, and predict the big winner.

In order to fill out tournament brackets with high predictive accuracy, many computer simulations and algorithms have been developed to model the tournament and attempt to explore the effective strategies for March Madness prediction. By reviewing the most successful modules (the winners of Kaggle’s previous years contests), I can define three major factors that are strongly influences the prediction accuracy - teams’ seeds, rating of teams learned from season games and team’s possession. On top of those factors I’ll add more variables from the Kaggle’s “March madness” data set and I’ll examine the possibility for external data.

On a personal note, beside dealing with one of my favorite sports, I hope to get more added value from the methods of this project that will be used by me on my daily work. Since I have a lot of performance data, and this kind of a performance-oriented prediction, should be beneficial for this goal.